Recent transcontinental sweep of *Toxoplasma gondii* driven by a single monomorphic chromosome

A. Khan[†], B. Fux[†], C. Su[‡], J. P. Dubey[§], M. L. Darde[¶], J. W. Ajioka[‖], B. M. Rosenthal[§††], and L. D. Sibley[†,††]

[†]Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63130; [‡]Department of Microbiology, University of Tennessee, Knoxville, TN 37996; [§]Animal Parasitic Disease Laboratory, Animal and Natural Resources Institute, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705; [¶]Faculté de Médecine, EA3174, Biological Resource Center for Toxoplasma, 87042 Limoges, France; and [‖]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, United Kingdom

Edited by John J. Mekalanos, Harvard Medical School, Boston, MA, and approved August 1, 2007 (received for review March 13, 2007)

Toxoplasma gondii **is a highly prevalent protozoan parasite that infects a wide range of animals and threatens human health by contaminating food and water. A markedly limited number of clonal parasite lineages have been recognized as predominating in North American and European populations, whereas strains from South America are comparatively diverse. Here, we show that strains from North America and Europe share distinct genetic polymorphisms that are mutually exclusive from polymorphisms in strains from the south. A striking exception to this geographic segregation is a monomorphic version of one chromosome (Chr1a) that characterizes virtually all northern and many southern isolates. Using a combination of molecular phylogenetic and phenotypic analyses, we conclude that northern and southern parasite populations diverged from a common ancestor in isolation over a period of $\approx 10^6$ yr, and that the monomorphic Chr1a has swept each population within the past 10,000 years. Like its definitive feline hosts, *T. gondii* may have entered South America and diversified there after reestablishment of the Panamanian land bridge. Since then, recombination has been an infrequent but important force in generating new *T. gondii* genotypes. Genes unique to a monomorphic version of a single parasite chromosome may have facilitated a recent population sweep of a limited number of highly successful *T. gondii* lineages.**

biogeography | evolution | pathogen | transmission | virulence

Approximately 25% of the world's human population is chronically infected by *Toxoplasma gondii*, which also parasitizes an exceptionally wide range of warm-blooded vertebrates (1). Its extraordinary prevalence depends on several adaptations for efficient transmission. Cats can excrete $>10^7$ environmentally resistant "oocysts" per day during an active infection (1), and these are capable of contaminating food or water. When ingested by intermediate hosts, the parasite differentiates to form tissue cysts containing semidormant parasites (bradyzoites) within muscle tissue and organs (1). The life cycle is completed when cats ingest such bradyzoites. Unlike most related coccidian parasites, bradyzoites of *T. gondii* are also infectious when ingested by other intermediate hosts (in which sexual recombination and oocyst formation do not occur) (1). Hence, *T. gondii* is unusual in that infection may propagate through sexual or entirely asexual means.

Although it possesses a meiotic life-cycle phase, *T. gondii* has a markedly clonal population structure, and three distinct lineages, types I, II, and III, together account for >95% of strains isolated in North America (NA) and Europe (E) (2–4). These three lineages are derived from only a few meiotic crosses among genetically very similar parental strains (5, 6). The limited genetic diversity within each of these lineages indicates they have arisen within the last ≈10,000 yr (7). When ingested, the bradyzoites of these clonal lineages all efficiently establish infection not only in cats but also in intermediate hosts, such as rodents (7). The broad host range of these clonal strains and their ability to perpetuate asexually among intermediate hosts may account for their disproportionate success and genetic cohesion (7). Strikingly, all three lineages share a nearly monomorphic version of one chromosome, Chr1a (referred to here as Mono-ChrIa), in stark contrast to the genetic variability characterizing all other chromosomes (8). Fixation of Mono-ChrIa in natural populations, despite the fact that this chromosome undergoes normal levels of recombination in experimental crosses (8, 9), is highly unexpected.

Although clonality clearly typifies isolates from NA and E, isolates from South America (SA) exhibit greater genetic diversity. When applied to strains from SA, markers designed to detect DNA polymorphisms in isolates from NA and E reveal patterns resembling mixtures of types I and III with a lower frequency of type II strains (10). Microsatelite analysis suggests four major lineages of *T. gondii*, two endemic to SA, one endemic in NA/E, and one occupying a global distribution (11). Although such methods detect genetic variation, they do not reveal all polymorphisms present in a given locus and may misclassify other variants due to homoplasy. Recently, intron sequencing has revealed otherwise previously unseen diversity in SA isolates and has demonstrated that they cannot be adequately characterized based solely on polymorphisms that discriminate among clonal lineages endemic to NA and E (12).

Here we establish the population structure of *T. gondii* by sequencing introns, which are presumed to be selectively neutral, and by examining the distribution and inheritance of Mono-ChrIa. Genetic lineages were also analyzed for traits of virulence and oral infectivity in mice. Our findings reveal both ancient and recent population changes that were shaped by biogeography over the past several million years and much more recently by the adaptive radiation of a small number of highly successful lineages.

Results

Phylogenetic Analysis Reveals a Marked Separation Between Strains Common to NA and E vs. SA. To define the divergence among *T. gondii* strains, we analyzed the frequency of SNPs within eight introns of five unlinked loci that collectively constituted 3,780 bp per strain [supporting information (SI) Table 1]. We compared exemplars of clonal groups I, II, and III with isolates previously considered "exotic" because of their dissimilarity from the clonal lineages and a select group of strains from SA

Author contributions: A.K., B.F., C.S., and L.D.S. designed research; A.K., B.F., and C.S. performed research; A.K., C.S., J.P.D., M.L.D., J.W.A., and B.M.R. contributed new reagents/analytic tools; A.K., J.W.A., B.M.R., and L.D.S. analyzed data; and B.M.R. and L.D.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Abbreviations: E, Europe; LD, linkage disequilibrium; MRCA, most recent common ancestry; NA, North America; SA, South America.

[††]To whom correspondence may be addressed. E-mail: benjamin.rosenthal@ars.usda.gov or sibley@wustl.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0702356104/DC1.

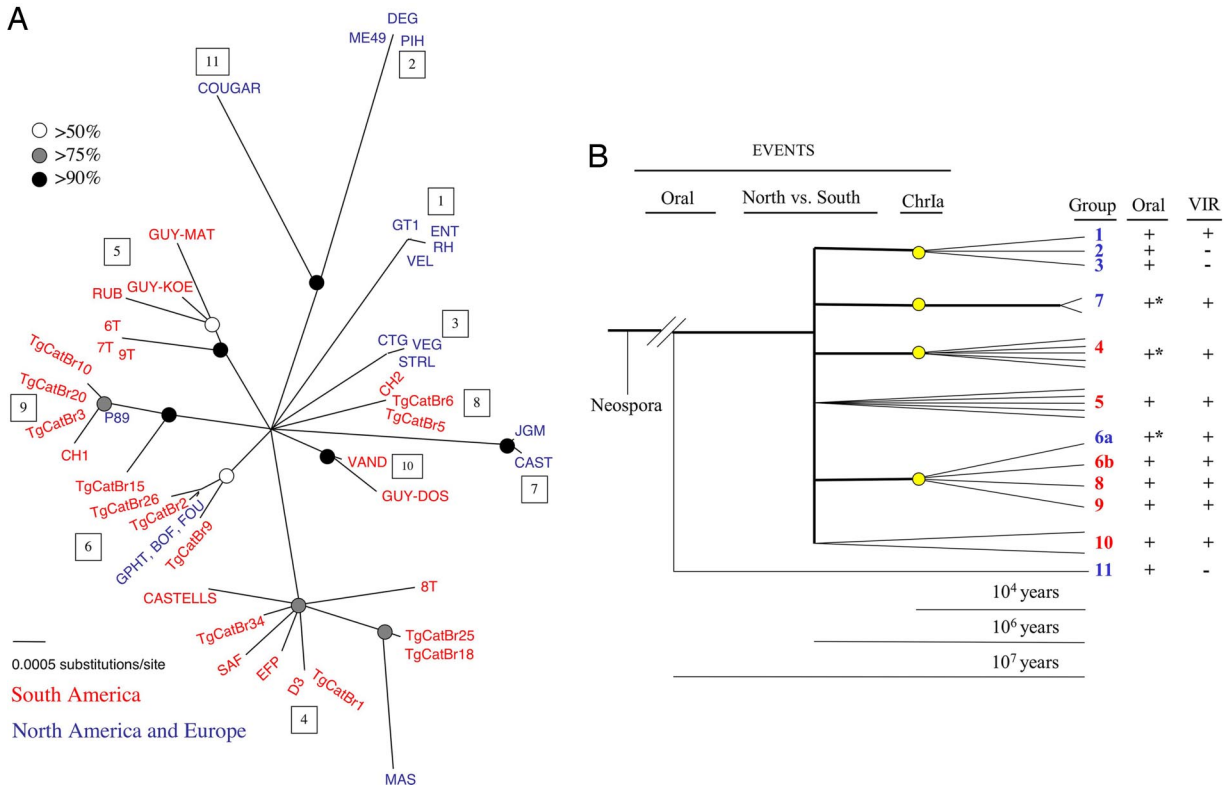

Fig. 1. *T. gondii* strains show marked geographic separation and clonality. (*A*) Phylogenetic analysis of *T. gondii* strains based on intron sequences identified 11 separate haplogroups (numbered in boxes), with striking geographic separation between NA and E (blue lettering) and SA (red lettering). Unrooted phylogram generated by using neighbor-joining analysis; bootstrap values are given by the percent at each node. (*B*) Diagram of the MRCA for strains of *T. gondii* (SI Table 5). The major North–South split occurred ≈10^6 yr ago; group 11 (COUG) precedes this split. The emergence of a monomorphic Chr1 in both NA/E and SA strains was ≈10,000 yr ago (small yellow circle). Phenotypic traits for oral transmission (ORAL) and acute virulence (VIR) are as described in SI Table 2.

(central and southern Brazil and French Guyana) (SI Table 2) (7). Neighbor-joining and parsimony analyses of the intron sequence data grouped the 46 different strains into 11 distinct haplogroups (Fig. 1*A*). The haplogroups occupy strikingly distinct geographic distributions; most strains from groups 1–3 (previously defined lineages I, II, and III) occurred almost exclusively in NA and E, whereas groups 4, 5, and 8–10 occurred primarily in SA, and group 6 was widespread, being found in E, SA, and Africa (3). Very few exceptions to this geographic separation were noted. The genetic makeups of NA and E populations were highly similar [pairwise fixation index (F_{ST}) = 0.0524, $P < 0.05$ overall and F_{ST} = 0.029, not significant, after removal of rare variants]. By contrast, SA isolates were strongly differentiated from both NA (pairwise F_{ST} = 0.211, $P < 0.0001$) and E (pairwise F_{ST} = 0.185, $P < 0.0001$). These analyses reaffirm that the same haplogroups predominate in NA and E, and that these differ markedly from those endemic to SA. Strains previously considered "exotic," turn out to be common lineages in SA, and appear unusual only by comparison to the well studied clonal lineages from NA and E.

Although sexual recombination would be expected to reassort allelic variation into many multilocus genotypes, the observed repertoire in NA and E is restricted to several lineages characterized by strong genome-wide linkage among physically unlinked alleles (SI Fig. 5). Certain multilocus genotypes endemic to SA were also repeatedly sampled, providing evidence that they also propagate clonally (Fig. 1*A* and SI Fig. 5). To examine linkage disequilibrium (LD) within and between loci, we joined the intron sequences end to end and analyzed pairwise associations of polymorphism across these sequences. LD was estimated by using the *Zns* statistic, which expresses the average correlation among alleles in all pairs of polymorphic sites (13). *Zns* was elevated not only for the clonal haplogroups 1, 2, and 3 from NA and E but also for SA groups 8, 9 ($P < 0.05$), and 5, 10 (SI Table 3). Other SA isolates (i.e., group 4) exhibited lower values of *Zns* (SI Table 3), as would be expected in lineages experiencing more frequent sexual recombination. A more conventional measure of LD, based on the D′ statistic, which expresses the difference between actual and expected dinucleotide haplotype frequencies under the null hypothesis of random mating, supported similar conclusions (SI Fig. 5). Remarkably, some lineages (i.e., groups 1, 2, and 3) showed complete LD even between unlinked loci, whereas others (i.e., groups 5, 8, 9, and 10) showed high LD within a given locus, with less association among unlinked loci (SI Fig. 5).

Distinct Biallelic Polymorphisms Have Accumulated in the North and South Over Approximately Equivalent Time Periods. Previous studies have noted that NA and E isolates from type I, II, and III (haplogroups 1, 2, and 3) are mixtures of biallelic polymorphisms that were inherited as large blocks across the genome, indicating they arose from only a few genetic crosses between highly similar parental strains (6). With the notable exception of Chr1a (discussed below), SNPs found in NA and E were fixed in SA isolates. Surprisingly, the converse is also true; SA strains also show striking biallelic haplotypes, yet these polymorphisms occurred at distinct positions from those seen in the NA/E strains (SI Table 4). Collectively, these patterns indicate that northern and southern strains have mutually exclusive bialellic polymorphic haplotypes.

The most parsimonious explanation of these data is that northern and southern strains have a common origin but have

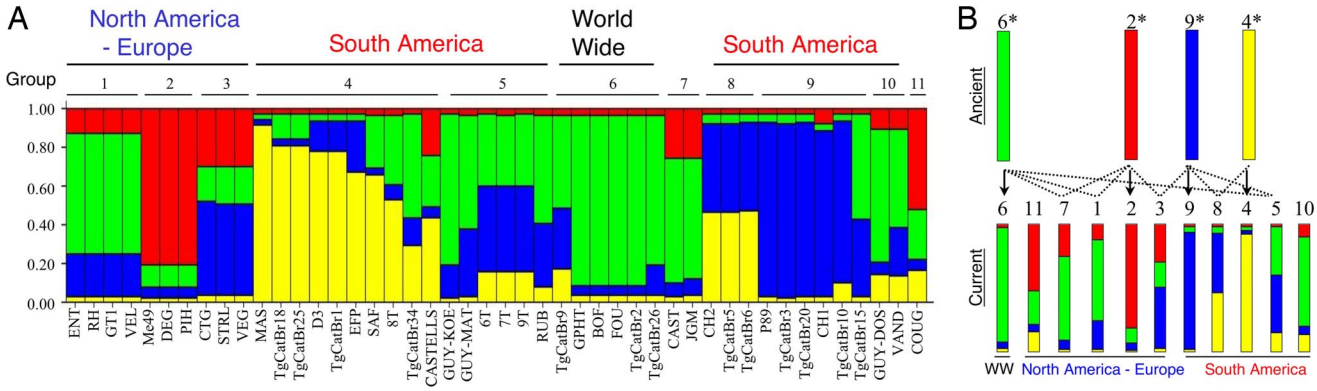

Fig. 2. Structure of *T. gondii* populations. (*A*) Analysis of present-day haplogroups using K = 4 populations identified 11 groups, corresponding to the haplogroups defined in Fig. 1. (*B*) Model for generation of current haplogroups by admixture of ancient groups resembling 2, 4, 6, and 9. WW, worldwide distribution.

accumulated separate characteristic mutations during an extended period of isolation. To test this model, we estimated the most recent common ancestry (MRCA) between strains from NA and E and those from SA based on the frequencies of SNPs in each population. Analysis of the frequency of biallelic polymorphism indicates that strains from both the North and South have a similar predicted MRCA of $\approx 10^6$ yr (SI Table 5) (Fig. 1*B*). The exception to this pattern was the strain COUG, which has a MRCA with other strains of close to 10^7 yr, indicating it diverged before the North–South split.

Recognizing that polymorphisms in clonal lineages are derived from a limited number of parental lineages, we then investigated the duration over which "new mutations" (i.e., those not already present in parental strains) have accumulated. Thus, we performed additional analyses excluding those polymorphisms exhibiting typical biallelic SNPs (but which, instead, appear sporadically in particular isolates). In both NA and SA, such clonal haplogroups coalesce ≈10,000 yr ago (SI Table 5) (Fig. 1*B*). Collectively, these patterns indicate that northern and southern populations diverged ≈1 Mya, and that much more recently, a small number of clonal groups have rapidly expanded within the past 10,000 yr.

Mixing of Four Ancestral Groups Can Explain the Current Population Structure. A Bayesian statistical approach was used to infer population structure from allelic variation in the intron sequences by using STRUCTURE (14). Because *T. gondii* appears to be composed of relatively few genotypes that have historically undergone limited but important admixture (6–8), we explored these data using a linkage model. The model most compatible with the current assemblage of strains suggests that they were derived from admixture of four ancestral lineages (SI Fig. 6). Eleven extant groups, corresponding closely to the haplogroups identified by phylogenic analysis (Fig. 1*A*), were identified using STRUCTURE (Fig. 2*A*). Each can be derived by limited admixture of the four inferred ancestral lineages, which most closely correspond to haplogroups 2, 4, 6, and 9 (Fig. 2*B*).

Most Strains of *T. gondii* Share a Monomorphic Chr1a. A recent comparison of whole-genome sequence of Chr1a and Chr1b from members of the three clonal lineages (1, 2, and 3) revealed they share a monomorphic version of Chr1a (8) (Mono-ChrIa). To determine how widespread this pattern might be, we sequenced 12 blocks scattered across Chr1a from 30 representative strains; these analyses revealed remarkably few polymorphisms for the majority of strains (Fig. 3*A*). Members of groups 4 (except strain CASTELLS), 7, 8, and 9 all contained a nearly identical Mono-Chr1a (Fig. 3*A*). Exceptions to this pattern included groups 5 and 10, which often contained a separate shared allele that differed from the Mono-ChrIa version by conserved biallelic polymorphisms ("Alternative"; Fig. 3*A*). Additionally, groups 5 and 10 also contained regions that differed substantially among each isolate ("Divergent"; Fig. 3*A*). Groups 6 and 11 had chimeric versions of ChrIa, in which approximately half of the chromosome was identical to Mono-Chr1a (Fig. 3*A*).

A neighbor-joining tree reconstructed from variation in the sequenced blocks depicts the divergence among strains based on Chr1a (SI Fig. 7). With the sole exception of COUG, strains in NA and E are characterized by Mono-Chr1a, as are isolates belonging to the SA haplotypes 4, 6b, 8, and 9. Notably, most groups characterized by Mono-ChrIa were also markedly clonal when analyzed at other loci in the genome (see Fig. 1 *A* and *B*). In contrast, groups 5, 10, and CASTELLS contained highly divergent versions of ChrIa (SI Fig. 7).

Inheritance of the Apicoplast Supports a Simple Recent Ancestry of Strains. Apicomplexans contain a 35-kb circular genome that is the remnant of a secondary endosymbiont: this organellar genome is inherited maternally (14) and therefore does not undergo genetic recombination. Coalescent analysis of polymorphisms in three regions of apicoplast genome sampled from 35 representative strains was used to infer the ancestral origin by using statistical parsimony (15). Network analysis of the apicoplast haplotypes showed a striking correlation between the Mono-ChrIa and inheritance from just a few matrilineages (Fig. 3*B*) (SI Table 6). Haplogroups 1, 2, 4, and 8 were derived from a single common matrilineage, whereas groups 3, 6, and 9 were descendant from a second distinct matrilineage (Fig. 3*B*). In contrast, strains that lacked Mono-ChrIa were spread across divergent nodes of the network (Fig. 3*B*). These results indicate that Mono-ChrIa may have arisen in a single genetic background and subsequently spread to the majority of haplogroups through very few genetic crosses.

Acute Virulence and Oral Transmission. The extremely widespread success of a small number of *T. gondii* lineages suggests they have a strong selective advantage. Previous studies suggested that improved transmissibility via oral transmission of tissue cysts between intermediate hosts may explain the predominance of types I, II, and III lineages (1, 2, and 3 here) (7). This trait favors clonal dissemination via carnivorous or omnivorous feeding between intermediate hosts. Additionally, recent studies have mapped genes responsible for acute virulence in the type I lineage in mice, and this trait may also constitute a selective advantage (16). We were therefore curious to understand how

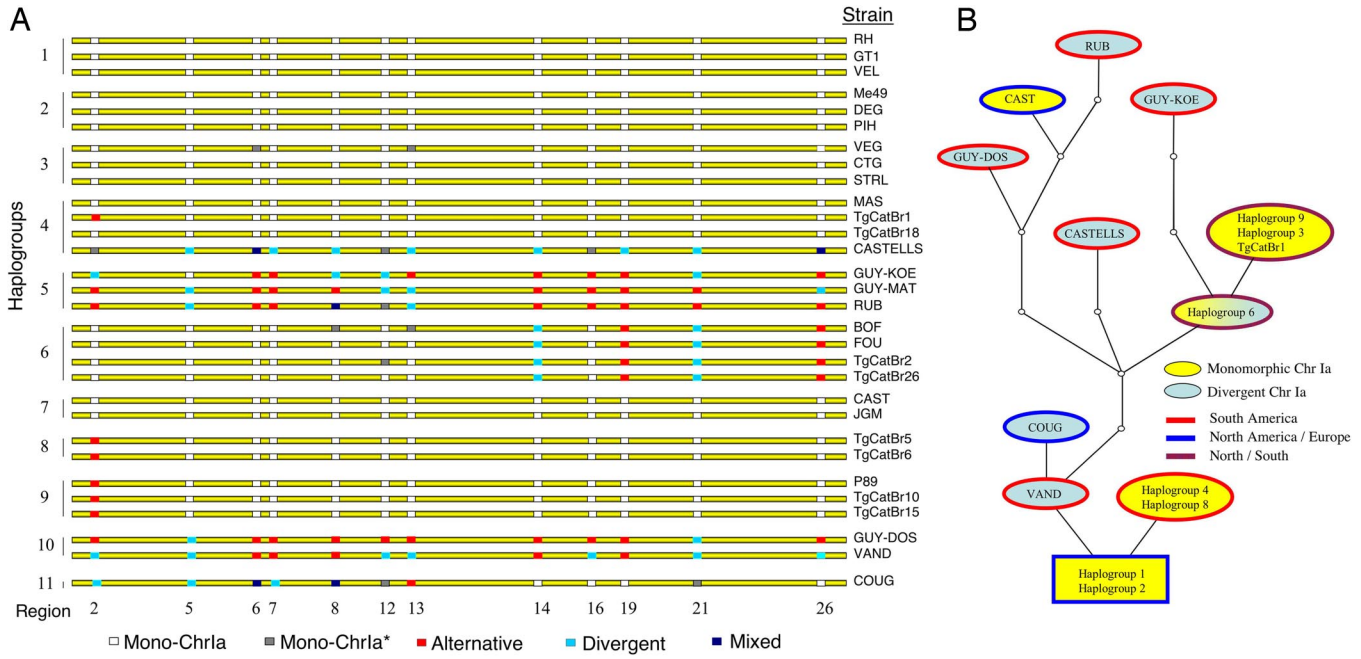

Fig. 3. Sequence divergence of ChrIa among strains of *T. gondii*. (*A*) Numbered regions (8) were sequenced from each of the stains shown. Mono-ChrIa (white) indicates identical. Alternative (red) indicates a separate shared allele. Divergent (blue) indicates highly distinct alleles. Mixed (green) represents recombinant regions. MonoChr1a* (gray) indicates minor variants (i.e., single SNP). Yellow regions are not sequenced. (*B*) Network analysis of apicoplast inheritance among select strains of *T. gondii*. Haplogroups with shared maternal inheritance typically also contained mono-ChrIa (yellow shading), whereas strains with divergent ChrIa (blue) fall on separate nodes of the network.

broadly such traits were distributed among strains representing greater geographic and genetic diversity.

Representative strains from each haplogroup (SI Table 2) were tested for oral transmissibility by feeding to mice tissue cysts that had developed in the brains of chronically infected mice, as defined previously (7). Efficient oral transmission proved to be a widespread trait in strains from all SA haplogroups, similar to the previously described clonal strains from NA and E (SI Table 2) (Fig. 1*B*). Rare exceptions occurred in specific strains from haplogroups 4, 6a, and 7 (* in Fig. 1*B*). In addition to confirming the lack of oral transmissibility of tissue cysts from several previously studied strains [i.e., CAST (group 7) and MAS (group 4)] (7), two additional isolates were thus identified [i.e., FOU (group 6a) and GPHT (group 6a)]. However, highly similar strains from groups 4, 6a, and 7 readily caused infection in mice fed tissue cysts (Fig. 1*B*) (SI Table 2). Our studies demonstrate that *T. gondii* tissue cysts of various ancestries (and not just those that have experienced clonal expansion) are infectious to mice upon oral ingestion, although mutation or genetic recombination may occasionally compromise tissue cyst development (17).

Type I strains are extremely virulent, and the effective LD_{100} is one viable organism (infection always leads to death) in outbred mice, whereas types II and III are relatively nonvirulent (4). We tested the acute virulence of select strains representing different haplogroups. Acute virulence, comparable to that of type I lineages, was widespread among SA haplogroups, including groups 4, 5, 6b, 8, and 9 (SI Table 2) (Fig. 1*B*). Given the ancient divergence of southern from northern strains, it appears, that acute virulence in mice is an ancestral trait, and it does not uniquely characterize recently evolved, clonally expanding *T. gondii* lineages.

Discussion

Previous studies have recognized a highly clonal population structure of *T. gondii* composed of three predominant lineages in NA and E, with rare variants being considered "exotic" (2, 3). Other studies have indicated that strains from SA were more genetically diverse, although their relationship to the clonal and exotic groups was unclear (11, 12). Here, we reconcile previous uncertainties about the population structure of *T. gondii* by demonstrating extreme geographic separation between strains that are predominant in NA and E vs. those that are common in several regions of SA. Overall, the population structure of *T. gondii* consists of discrete lineages that tend to be highly clonal and that show strong geographic restriction. Many of these clonal lineages occupy broad geographic ranges where they are found in multiple hosts. These data are most consistent with a model that strains in NA, E, and SA have a common ancestry, but that they have diverged and evolved in two distinct phases: (*i*) differentiation of northern and southern haplogroups occurred among geographically isolated populations, and (*ii*) a recent global sweep has homogenized many regions, driven by inheritance of a monomorphic version of ChrIa (Fig. 4).

Our analyses confirm that SA strains embody genetic diversity absent in NA and E (11) and extend this insight by recognizing 11 distinct haplogroups of *T. gondii*, four of which are almost exclusively found in SA (haplogroups 4, 5, 8, and 9), and one of which is globally distributed (haplogroup 6). Previously characterized "exotic" strains are now recognized as belonging to one of the common haplogroups, albeit existing outside their normal range. *T. gondii* infects a wide range of warm-blooded vertebrates, including birds, many of which are migratory. Hence it is unexpected that greater mixing has not occurred between continents. Examination of *T. gondii* strains from other regions of South and Central America may reveal evidence of such exchange, but are unlikely to alter the basic North vs. South dichotomy revealed here. Recognizing that most SA strains are demarcated by an entirely distinct array of polymorphic haplotypes, it is no longer appropriate to type SA strains using those markers that were designed to discriminate among strains in NA and E.

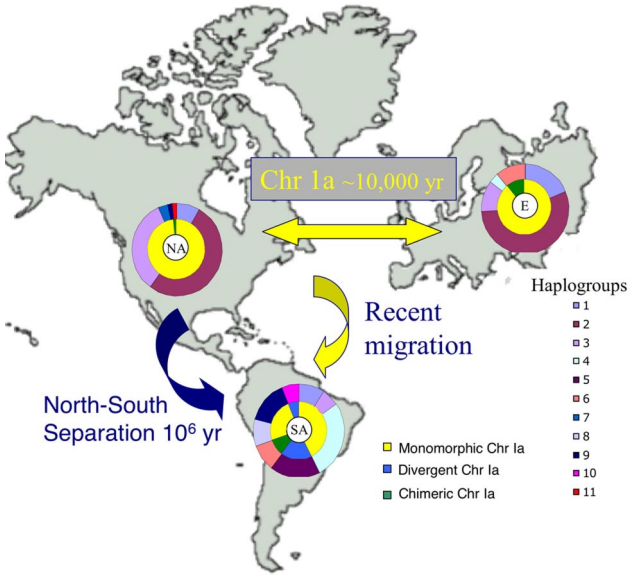

Fig. 4. Proposed model for the dissemination of *T. gondii* strains in E, NA, and SA. The spread from NA to SA is estimated at 10^6 yr (curved blue arrow). Emergence of Mono-Chr1a occurred ≈10,000 yr ago (horizontal yellow arrow) and spread between NA and E. More recently, Mono-Chr1a penetrated into SA (curved yellow arrow). The outer color wheels show the prevalence of haplogroups in current populations; the inner color wheel depicts ChrIa. Strains used include those studied here plus previously isolated strains (2, 12, 19, 27).

The finding of increased genetic diversity among SA strains using intron variation mirrors similar conclusions derived from microsatelite variation (11); however, we find little evidence for the suggestion that strains from SA are older. Rather, coalescent analysis indicates that variation in NA/E vs. SA appears to have accumulated over approximately equivalent time. Interestingly, the initial divergence between NA and SA populations of *T. gondii* coincides with the reconnection of the Panamanian land bridge (1–2 Mya during the upper Pliocene) after ≈50 Mya of separation (dating back to the late Eocene), an event that facilitated the southerly migration (and subsequent speciation) of its definitive hosts, members of the cat family (Felidae) (18). Comigration of the parasite with its definitive host may therefore have led to the divergence of southern strains of *T. gondii* from preexisting northern ones (Fig. 4). Consistent with this, the genetic diversity of *T. gondii* is greater in the South, similar to the situation observed in felids (18).

Although recombination has generated a greater diversity of *T. gondii* genotypes in SA than elsewhere, clonal propagation also characterizes some SA haplogroups. Certain genotypes are repeatedly recovered from epidemiologically unrelated SA samples, and many others manifest strong LD in physically unlinked locus pairs and are tightly clustered in star-shaped phylogenies. Our analyses suggest that ancient population patterns in the North, and to a lesser extent in the South, have been substantially obscured by the comparatively recent emergence of a small number of clonal strains. The greater genetic diversity in the South, which has led others to conclude that lineages from this region are older (11), may instead reflect the incomplete penetrance of Mono-ChrIa in this region (Fig. 4).

Here, we extend to additional variants the earlier observation that genetic complexity among clonal lineages of *T. gondii* in NA and E can be explained by just a few genetic crosses between similar strains in the wild (6). Two unique parental strains, when separately crossed with a type II progenitor, likely gave rise to types I and III, respectively (6). The unique parental strain for type III was recognized to strongly resemble P89, then considered an "exotic strain" (6). Here, we demonstrate that P89 belongs to a prevalent SA haplogroup (group 9), which is highly similar to one of the four ancestral groups (Fig. 2*B*). The parental strain for type I was previously unknown, but our studies suggest it was derived from a progenitor that resembled haplogroup 6, another major ancestral group (Fig. 2*B*). Our findings indicate that limited genetic exchange among four ancestral populations may have given rise to all of the currently recognized extant haplogroups. The ancestral population was likely composed of lineages resembling haplogroups endemic to the North (haplogroup 2), the South (haplogroups 4, 9), and a pandemic group (haplogroup 6). Albeit infrequent, sexual recombination appears to have played an extremely important role in shaping the population structure of *T. gondii*.

Remarkably, clonally perpetuating strains in SA harbor the same invariant version of Chr1a (Mono-ChrIa) that characterizes the three clonal lineages in the North (8). It is striking that an entire chromosomal variant, estimated to have originated within the last 10,000 yr (8), has become established in genomes otherwise estimated to have last shared common ancestry 1 Mya. Collectively, these findings indicate a recent origin and rapid spread of Mono-ChrIa associated with the expansion of highly clonal lineages that have emerged over the last 10,000 yr (Fig. 4). The penetrance of Mono-ChrIa into other regions of the world (i.e., other regions of SA, Asia, and Africa) is presently unknown, because relatively few studies have explored the genetic diversity of *T. gondii* strains from these regions. Future studies need to adequately address genetic diversity by direct sequencing of selectively neutral loci and in particular assess the genetic homogeneity of ChrIa from broader regions of the world.

The extreme success of lineages that harbor Mono-ChrIa, which has rapidly spread through a wide range of hosts across much of three major continents, suggests it carries some major selective advantage. Several traits could potentially explain this, including: (*i*) enhanced oral transmission between intermediate hosts, thus bypassing the need for sexual reproduction; (*ii*) enhanced virulence leading to higher tissue burdens and hence greater transmission; and (*iii*) especially high reproductive fecundity in the cat, such that other stains are effectively outcompeted. Although Mono-ChrIa does not show a strong association with enhanced oral transmission or virulence in the laboratory mouse model, this highly permissive host may not best predict those phenotypes most important for transmission in the wild. Oral transmissibility of tissue cysts to intermediate hosts differentiates *T. gondii* from many, if not all, of its closest coccidian relatives, and its highly clonal population structure argues that such strictly asexual transmission contributes markedly to its exceptionally widespread dissemination.

Strains of *T. gondii* differ phenotypically despite their generally limited genetic variability, and certain genotypes cause especially severe human disease. Type I strains are especially prevalent in some studies of AIDS patients (19), highly divergent strains have been shown to present with atypical clinical severity (20), and certain SA strains have been associated with severe ocular disease (12). Understanding why Mono-ChrIa has swept disparate *T. gondii* populations may elucidate how genes responsible for virulence have spread through this parasite's population. Comparing the content and expression of its ≈120 genes (8) in parasites that harbor the Mono-ChrIa to those containing divergent versions of Chr1a may identify the selective basis for its recent sweep of *T. gondii* populations.

Materials and Methods

Analysis of DNA Polymorphisms. Sequencing of introns was conducted on PCR-amplified templates from 46 representative strains by using BigDye cycle sequencing (Applied Biosystems, Foster City, CA) (performed by SeqWright DNA Technology Services, Houston, TX). Primers used for sequencing are listed in SI Table 1;

primary data are summarized at www.toxohapmap.wustl.edu. Clustal (21) was used to align the sequences that were analyzed by using PAUP*4.0B (22). Unrooted phylogenetic comparisons were conducted with distance and parsimony methods. The conditions were set to distance (mean character difference, minimum evolution, negative branches = 0), and 1,000 bootstrap replicates were performed by using the BioNeighbor-Joining algorithm. Alternatively, parsimony analysis was conducted by heuristic stepwise searching, with bootstrapping for >1,000 replicates. Consensus trees were drawn with an arbitrary root according to the bootstrap 50% majority rule. Pairwise F_{ST} values between NA, E, and SA *T. gondii* strains were calculated in Arlequin Ver 3.01 with 10,000 permutations (23).

LD. Intron sequences from different loci were combined in a single file and analyzed for LD across the entire composite sequence. Haplotype blocks and LD plots were constructed by using DnaSP 4.0 (24). D' was calculated for all pairs of sites. Both the two-tailed Fisher's exact test and the χ^2 test were computed to determine significance of the associations between polymorphic sites. The average LD was estimated by using the *Zns* statistic (13), which averages LD over all pairwise comparisons for S polymorphisms in N sequences.

Analysis of Populations. Genotype data for unlinked markers (introns) were used to infer ancestral population patterns by using the program STRUCTURE 2.0 (25) run under the assumption of linkage with independent allele frequencies. Population sizes (K) of 2–10 were tested by five runs each using 10^5 burn in iterations followed by 10^6 iterations using the Markov Chain Monte Carlo method. The highest-probability runs were selected and compared for log likelihood and lambda, the allele frequency, to estimate the most probable model.

Sequence Analysis of Chr1a Blocks. Twelve scattered blocks of 800–900 bp were chosen from a group of 26 previously characterized regions of ChrIa (8). Each region was amplified with gene-specific primers and sequenced from 30 representative strains (primary data are summarized at: www.toxohapmap.wustl.edu). Sequences were aligned in Clustal and classified based on SNP profile.

Analysis of Apicoplast Regions. Sequences were determined from three loci within the circular apicoplast genome (NC_001799) from the segments called AP1, AP2, and AP3 for a total of 1,685 bp per strain for 35 strains. Primers used for sequencing are listed in SI Table 1; primary data are summarized at www.toxohapmap.wustl.edu. Sequences were aligned by using Clustal, and the output nexus file was analyzed by using TCS 1.21 (26) to determine the probability of parsimony for pairwise differences. A graphical output file of the resulting network was generated with missing intermediates and displayed by using the VGJ 1.0.3 drawing tool (http://www.eng.auburn.edu/department/cse/research/graph_drawing/graph_drawing.html).

Age Calculations. Estimates of the time to MRCA were measured from the rates of polymorphism between strains by using several estimates of the neutral mutation rate from the closely related parasite *Plasmodium falciparum*, as described (7). For estimating the common ancestry between northern and southern strains, we compared all strains by using the biallelic polymorphisms present in each of these regions. For estimating the ancestry of the clonal lineages, we pooled and compared those haplogroups that were dominant in the North vs. those in the South and removed the ancestral biallelic SNPs.

Animal Infections. Virulence of *T. gondii* strains was monitored after i.p. inoculation of parasites grown in culture as described (16). Oral transmission of bradyzoites from tissue cysts was monitored as described (7). Further details are found in SI Table 2.

We are grateful to Henry Huang (Washington University), John Wootton (National Center for Biotechnology Information, Bethesda, MD), Steve Beverley (Washington University), Rubens Belfort, Jr. (Paulista School of Medicine, Sâo Paulo, Brazil), Jack Remington (Palo Alto Medical Research Foundation, Palo Alto, CA), Jack Frenkel (Sante Fe, NM), and Ricardo Vitor (Fed. Univ. Minas Gerais, Belo Horizonte, Brazil) for strains of *T. gondii* and to Robert Cole (Washington University) and Julie Nawas (Washington University) for expert technical assistance. This work was supported by National Institutes of Health Grants AI036629 and AI059176 (to L.D.S.), the Biotechnology and Biological Sciences Research Council, the Wellcome Trust (J.W.A.), U.S. Department of Agriculture CRIS Project 1265-42000-010 (to B.M.R.), and the Department of Microbiology, University of Tennessee (C.S.).

1. Joynson DH, Wreghitt TJ (2001) *Toxoplasmosis: A Comprehensive Clinical Guide* (Cambridge Univ Press, Cambridge, UK).
2. Howe DK, Sibley LD (1995) *J Infect Dis* 172:1561–1566.
3. Ajzenberg D, Bañuls AL, Su C, Dumètre A, Demar M, Carme B, Dardé ML (2004) *Intl J Parasitol* 34:1185–1196.
4. Sibley LD, Boothroyd JC (1992) *Nature* 359:82–85.
5. Grigg ME, Bonnefoy S, Hehl AB, Suzuki Y, Boothroyd JC (2001) *Science* 294:161–165.
6. Boyle JP, Rajasekar B, Saeij JPJ, Ajioka JW, Berriman M, Paulsen I, Sibley LD, White M, Boothroyd JC (2006) *Proc Natl Acad Sci USA* 103:10514–10519.
7. Su C, Evans D, Cole RH, Kissinger JC, Ajioka JW, Sibley LD (2003) *Science* 299:414–416.
8. Khan A, Bohme U, Kelly KA, Adlem E, Brooks K, Simmonds M, Mungall K, Quail MA, Arrowsmith C, Chillingworth T, Churcher C, *et al.* (2006) *Gen Res* 16:1119–1125.
9. Khan A, Taylor S, Su C, Mackey AJ, Boyle J, Cole RH, Glover D, Tang K, Paulsen I, Berriman M, *et al.* (2005) *Nucleic Acids Res* 33:2980–2992.
10. Ferreira AM, Vitor RW, Gazzinelli RT, Melo MN (2006) *Infect Genet Evol* 6:22–31.
11. Lehmann T, Marcet PL, Graham DH, Dahl ER, Dubey JP (2006) *Proc Natl Acad Sci USA* 103, 11423–11428.
12. Khan A, Jordan C, Muccioli C, Vallochi AL, Rizzo LV, Belfort R, Jr, Vitor RW, Silveira C, Sibley LD (2006) *Emerg Infect Dis* 12:942–949.
13. Kelly JK (1997) *Genetics* 146:1197–1206.
14. Ferguson DJ, Henriquez FL, Kirisits MJ, Muench SP, Prigge ST, Rice DW, Roberts CW, McLeod RL (2005) *Eukaryot Cell* 4:814–826.
15. Templeton AR, Crandall KA, Sing CF (1992) *Genetics* 132:619–633.
16. Taylor S, Barragan A, Su C, Fux B, Fentress SJ, Tang K, Beatty WL, Haijj EL, Jerome M, Behnke MS, *et al.* (2006) *Science* 314:1776–1780.
17. Fux B, Suetterlin J, Khan A, Gill D, Su C, Sibley LD (2007) *Infect Immun* 75:2580–2590.
18. Johnson WE, Eizirik E, Pecon-Slattery J, Murphy WJ, Antunes A, Teeling E, O'Brien SJ (2006) *Science* 311:73–77.
19. Khan A, Su C, German M, Storch GA, Clifford D, Sibley LD (2005) *J Clin Microbiol* 43:5881–5887.
20. Carme B, Bissuel F, Ajzenberg D, Bouyne R, Aznar C, Demar M, Bichat S, Louvel D, Bourbigot AM, Peneau C, *et al.* (2002) *J Clin Microbiol* 40:4037–4044.
21. Higgins DG, Thompson JD, Gibson TJ (1996) *Methods Enzymol* 266:382–402.
22. Swofford DL (2002) *Phylogenetic Analysis Using Parsimony (* and Other Methods)* (Sinauer, Sunderland, MA).
23. Excoffier L, Laval G, Schneider S (2005) *Evol Bioinform Online* 1:47–50.
24. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) *Bioinformatics* 19:2496–2497.
25. Falush D, Stephens M, Pritchard JK (2003) *Genetics* 164:1567–1587.
26. Clement M, Posada D, Crandall KA (2000) *Mol Ecol* 9:1657–1659.
27. Howe DK, Honoré S, Derouin F, Sibley LD (1997) *J Clin Microbiol* 35:1411–1414.